

# Genome Browsers

**Shamith Samarajiwa**

Integrative Systems Biomedicine Group  
MRC Cancer Unit  
University of Cambridge

22th Sept 2015,  
CRUK Cambridge Institute



UNIVERSITY OF  
CAMBRIDGE



CANCER  
RESEARCH  
UK

CAMBRIDGE  
INSTITUTE

# Genomic Coordinate Systems

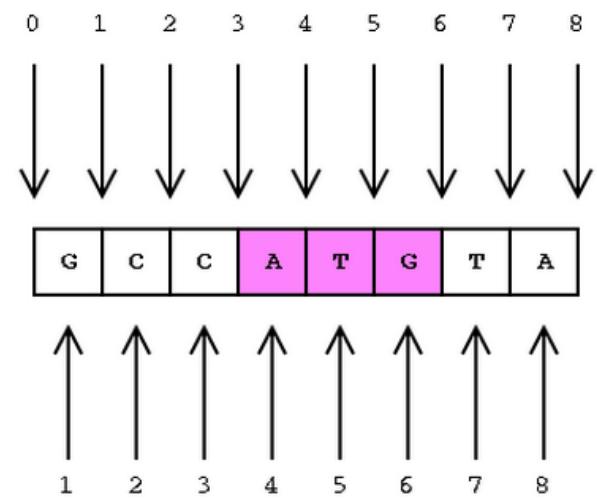
- There are two major coordinate systems in genomics.
- **Base coordinate system** anchors genomic feature to nucleotide positions while the **Interbase coordinate system** anchor genomic feature between nucleotide positions.
- Most genome annotation portals (e.g. **NCBI or Ensembl**), bioinformatics software (e.g. BLAST) and annotation file formats (e.g. GFF) use the base coordinate system, which represents a feature starting at the first nucleotide as **position 1**.
- Other systems (e.g. **UCSC, Chado, DAS2**) use the interbase coordinate system, whereby a feature starting at the first nucleotide is represented as **position 0**.

# Genomic Coordinate Systems

- The UCSC genome browser uses both systems and refer to the base coordinate system as “**one-based, fully-closed**” (used in the UCSC genome browser display) and interbase coordinate system as “**zero-based, half-open**” (used in their tools and file formats).
- The interbase coordinate system is also referred to as “space-based” by some authors.

There are several advantage for using the interbase coordinate system including:

1. the ability to represent features that occur between nucleotides (like a splice site),
2. simpler arithmetic for computing the length of features ( $\text{length}=\text{end}-\text{start}$ ) and overlaps ( $\max(\text{start1}, \text{start2}), \min(\text{end1}, \text{end2})$ )
3. more rational conversion of coordinates from the positive to the negative strand



# UCSC genome browser: Introduction

main sections:

1. UCSC Genome Browser
  2. BLAT
  3. Custom tracks, Sessions and Track Hubs
  4. Table Browser
  5. Other UCSC tools
- 
- what does it do?
  - How do I use it?
  - What problems does it help me solve?

## About the UCSC Genome Bioinformatics Site

Welcome to the UCSC Genome Browser website. This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides portals to [ENCODE](#) data at UCSC (2003 to 2012) and to the [Neandertal](#) project. Download or purchase the Genome Browser source code, or the Genome Browser in a Box ([GBIB](#)) at our [online store](#).

We encourage you to explore these sequences with our tools. The [Genome Browser](#) zooms and scrolls over chromosomes, showing the work of annotators worldwide. The [Gene Sorter](#) shows expression, homology and other information on groups of genes that can be related in many ways. [Blat](#) quickly maps your sequence to the genome. The [Table Browser](#) provides convenient access to the underlying database. [VisiGene](#) lets you browse through a large collection of *in situ* mouse and frog images to examine expression patterns. [Genome Graphs](#) allows you to upload and display genome-wide data sets.

The UCSC Genome Browser is developed and maintained by the Genome Bioinformatics Group, a cross-departmental team within the [UC Santa Cruz Genomics Institute](#) at the University of California Santa Cruz ([UCSC](#)). If you have feedback or questions concerning the tools or data on this website, feel free to contact us on our [public mailing list](#).

The Genome Browser project team relies on public funding to support our work. Donations are welcome -- we have many more ideas than our funding supports! If you have ideas, drop a comment in our [suggestion box](#).

[DONATE NOW](#)

### News

[News Archives ▶](#)

To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the [genome-announce](#) mailing list. Please see our [blog](#) for posts about Genome Browser tools, features, projects and more.

#### 29 June 2015 - GENCODE Genes Now the Default Gene Set on the Human (GRCh38/hg38) Assembly

In a move towards standardizing on a common gene set within the bioinformatics community, UCSC has made the decision to adopt the GENCODE set of gene models as our default gene set on the human genome assembly. Today we have released the GENCODE v22 comprehensive gene set as our default gene set on human genome assembly GRCh38 (hg38), replacing the previous default UCSC Genes set generated by UCSC. To facilitate this transition, the new gene set employs the same familiar UCSC Genes schema, using nearly all the same table names and fields that have appeared in earlier versions of the UCSC set.

By default, the browser displays only the transcripts tagged as "basic" by the GENCODE Consortium. These may be found in the track labeled "GENCODE Basic" in the Genes and Gene Predictions track group. However, all the transcripts in the GENCODE comprehensive set are present in the tables, and may be viewed by adjusting the track configuration settings for the All GENCODE super-track. The most recent version of the UCSC-generated genes can still be accessed in the track "Old UCSC Genes".

The new release has 195,178 total transcripts, compared with 104,178 in the previous version. The total number of canonical genes has increased from 48,424 to 49,534. Comparing the new gene set with the previous version:

- 9,459 transcripts did not change.
- 22,088 transcripts were not carried forward to the new version.
- 43,681 transcripts are "compatible" with those in the previous set, meaning that the two transcripts show consistent splicing. In most cases, the old and new transcripts differ in the lengths of their UTRs.
- 28,950 transcripts overlap with those in the previous set, but do not show consistent splicing (i.e., they contain overlapping introns with differing splice sites)

More details about the new GENCODE Basic track can be found on the [GENCODE Basic track description page](#).

#### 26 June 2015 - New Bonobo (panPan1) Assembly Now Available in the Genome Browser

We are pleased to announce the release of a Genome Browser for the May 2012 assembly of bonobo, *Pan paniscus* (Max-Planck Institute panPan1, UCSC version panPan1). The assembly was provided by the [Max-Planck Institute for Evolutionary Anthropology](#). There are 10,867 scaffolds with a total size of 2,869,190,071 bases.

Bulk downloads of the sequence and annotation data are available via the Genome Browser [FTP server](#) or the [Downloads](#) page. These data have [specific conditions for use](#). The bonobo (panPan1) browser annotation tracks were generated by UCSC and collaborators worldwide. See the [Credits](#) page for a detailed list of the organizations and individuals who contributed to this release.

**12 June 2015 - Data Integrator:** Have you ever wished that the Table Browser could associate your custom track items with some other track, while retaining the item names from both? We have released a new tool that can do just that, and more: the Data Integrator. [Read more](#).

**28 May 2015 - New UCSC Genes Track Released for GRCm38/mm10:** We're happy to announce the release of an updated UCSC Genes track for the GRCm38/mm10 mouse Genome Browser. [Read more](#).

# UCSC Genome Bioinformatics



CSHL Press | Journal Home | Subscriptions | eTOC Alerts | BioSupplyNet

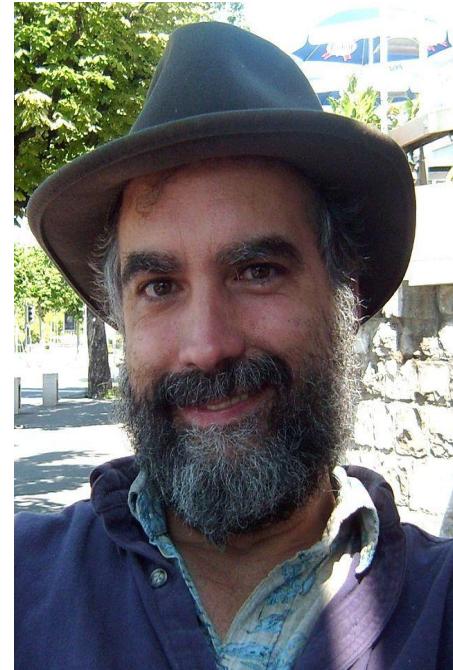
Genome Res. 2001 Sep; 11(9): 1541–1548.

doi: [10.1101/gr.183201](https://doi.org/10.1101/gr.183201)

## Assembly of the Working Draft of the Human Genome with GigAssembler

W. James Kent<sup>1,3</sup> and David Haussler<sup>2</sup>

**David Haussler**



**Jim Kent**

# 1. UCSC Browser

- Understanding the browser interface
- Basic searches
- Viewing tracks
- Configuring the display
- Navigating
- Printing images
- Retrieving DNA sequences and annotation

# Graphical view of genes, gene structure and annotation

Genomes   Genome Browser   Tools   Mirrors   Downloads   My Data   View   Help   About Us

## UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> zoom in 1.5x 3x 10x base ZOOM out 1.5x 3x 10x 100x

chr9:21,076,124-21,078,923 2,800 bp. enter position, gene symbol or search terms go

chr9 (p21.3) 24.2 9p24.1 9q23 22.3 p21.3 21.2 9p21.1 p13.3 13.1 9q12 9q11.2 9q12 9q13 9q11.1 9q21.13 21.31 p21.33 22.33 9q31.1 q31.2 q9q31.3 9q32 9q33.1 q33.2 q9q33.3 34.11 9q34.3

UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)

IFNB1

move start < 2.0 > move end < 2.0 >

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

track search default tracks default order hide all add custom tracks track hubs configure reverse resize refresh

collapse all expand all

Use drop-down controls below and press refresh to alter tracks displayed.

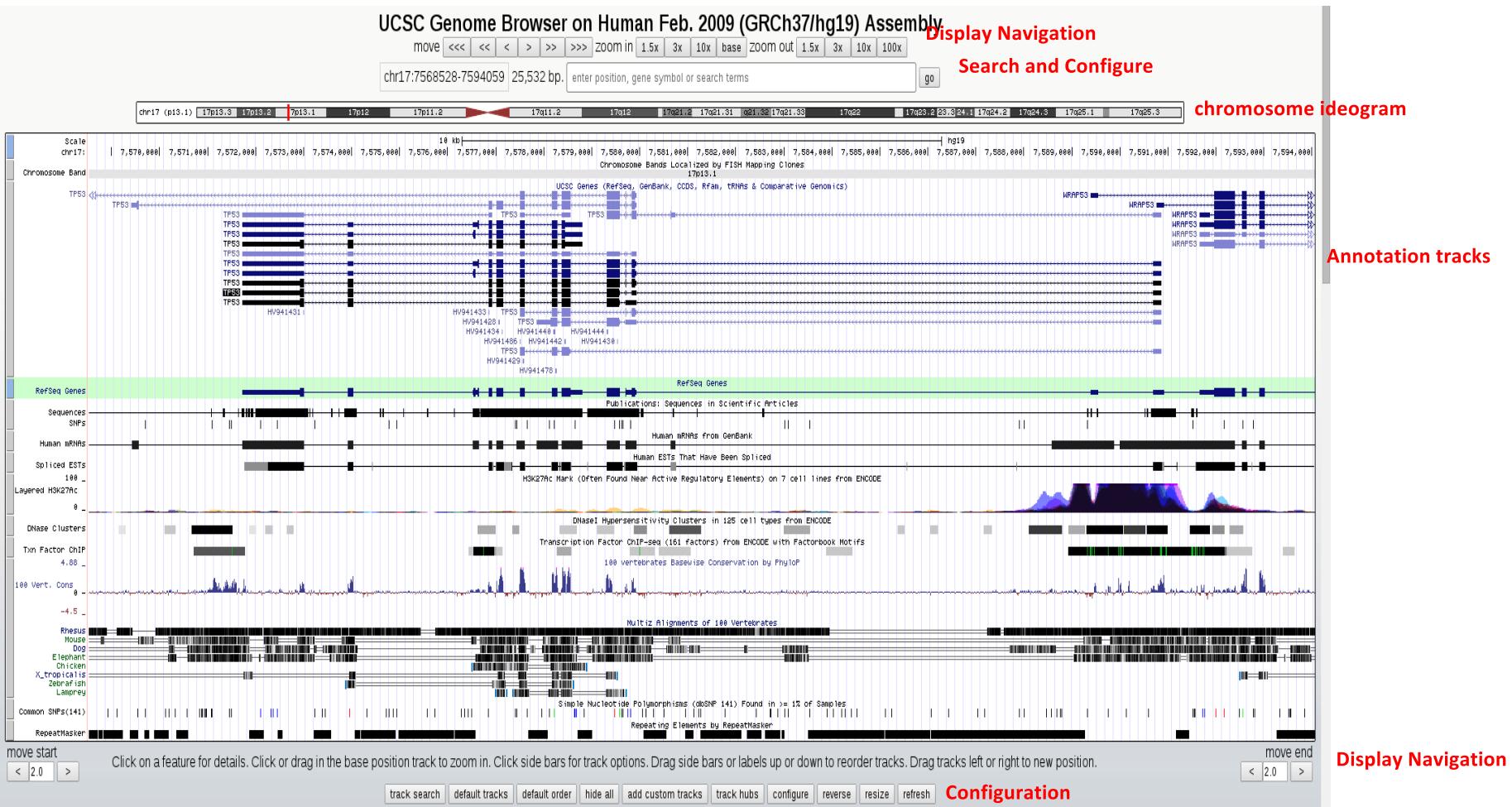
Tracks with lots of items will automatically be displayed in more compact modes.

Annotation

- + Mapping and Sequencing refresh
- + Genes and Gene Predictions refresh
- + Phenotype and Literature refresh
- + mRNA and EST refresh
- + Expression refresh
- + Regulation refresh
- + Comparative Genomics refresh
- + Neandertal Assembly and Analysis refresh
- + Denisova Assembly and Analysis refresh
- + Variation refresh
- + Repeats refresh

refresh

# Browser Interface



# Track Configuration

- Track configuration depends on track type and enables you to;
  - Set data thresholds
  - Include or exclude data from a specific source
  - Choose data labels
  - Choose graph type, height, range and scale
- Track and element descriptions contain additional information

# Configuring the genome browser display

track search default tracks default order hide all add custom tracks track hubs configure reverse resize refresh

**Search Advanced**

e2f1

search clear cancel

return to browser (0 of 10 selected)

Visibility	Track Name	Sort: <input checked="" type="radio"/> by Relevance <input type="radio"/> Alphabetically <input type="radio"/> by Hierarchy
hide ▾	<a href="#">HeLa E2F1 Std</a>	HeLa-S3 E2F1 Standard ChIP-seq Signal from ENCODE/SYDH ▾
hide ▾	<a href="#">HeLa E2F1 Std</a>	HeLa-S3 E2F1 Standard ChIP-seq Peaks from ENCODE/SYDH ▾
hide ▾	<a href="#">MCF-7 E2F1</a>	MCF-7 TFBS Uniform Peaks of HA-E2F1 from ENCODE/USC/Analysis ▾
hide ▾	<a href="#">HeLa-S3 E2F1 c2</a>	HeLa-S3 TFBS Uniform Peaks of HA-E2F1 from ENCODE/USC/Analysis ▾
hide ▾	<a href="#">HeLa-S3 E2F1 c1</a>	HeLa-S3 TFBS Uniform Peaks of E2F1 from ENCODE/USC/Analysis ▾
hide ▾	<a href="#">MCF7 HAE2 UCD</a>	MCF-7 HA-E2F1 UC Davis ChIP-seq Signal from ENCODE/SYDH ▾
hide ▾	<a href="#">MCF7 HAE2 UCD</a>	MCF-7 HA-E2F1 UC Davis ChIP-seq Peaks from ENCODE/SYDH ▾
hide ▾	<a href="#">HeLa HAE2 Std</a>	HeLa-S3 HA-E2F1 Standard ChIP-seq Signal from ENCODE/SYDH ▾
hide ▾	<a href="#">HeLa HAE2 Std</a>	HeLa-S3 HA-E2F1 Standard ChIP-seq Peaks from ENCODE/SYDH ▾
hide ▾	 <a href="#">SYDH TFBS</a>	Transcription Factor Binding Sites by ChIP-seq from ENCODE/Stanford/Yale/USC/Harvard ▾

Search for data types

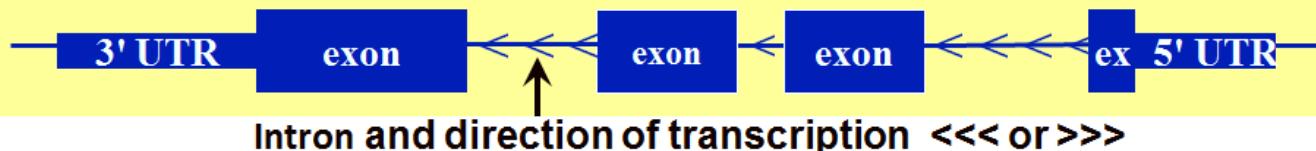
Return to Browser (0 of 10 selected)

 Tracks so marked are containers which group related data tracks. Containers may need additional configuration (by clicking on the  icon) before they can be viewed in the browser.

# Visual cues

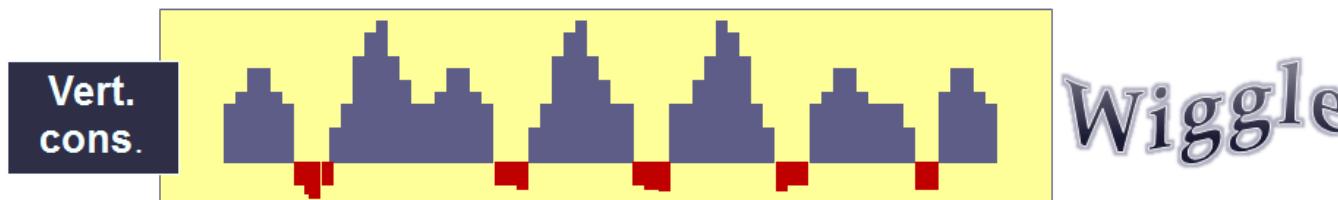


Tick marks; a single location (STS, SNP)



Track colors may have meaning—for example, UCSC Gene track:

- If there is a corresponding PDB entry = black
- If there is a corresponding reviewed/validated seq = dark blue
- If there is a non-RefSeq seq = lightest blue



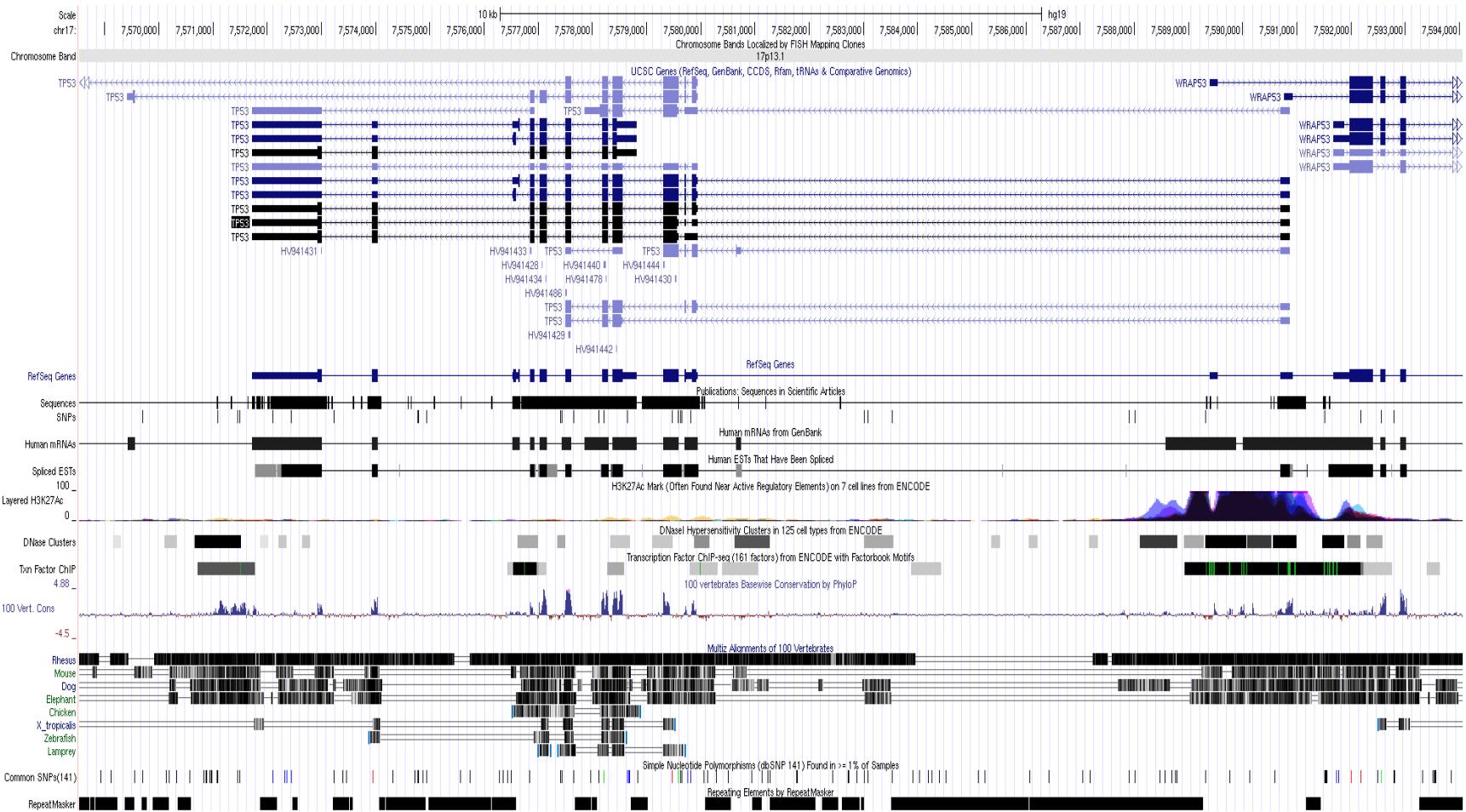
height of a blue bar is increased likelihood of conservation,  
red indicates a likelihood of faster-evolving regions

Alignment indications (Conservation pairs: “chain” or “net” style)

- Alignments = boxes, Gaps = lines



# Example search for human TP53





# Basic Annotation Track menu options

- Hide: removes a track from view



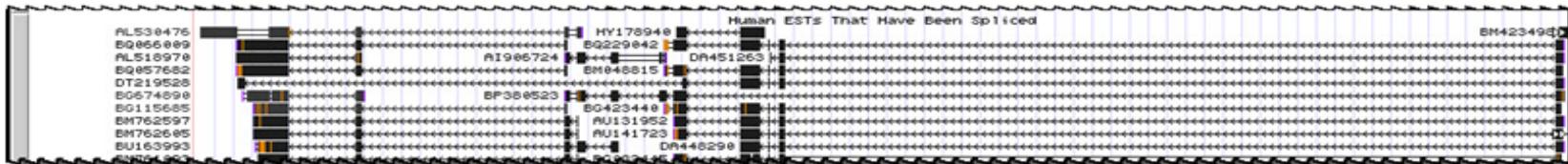
- Dense: all items collapsed into a single line



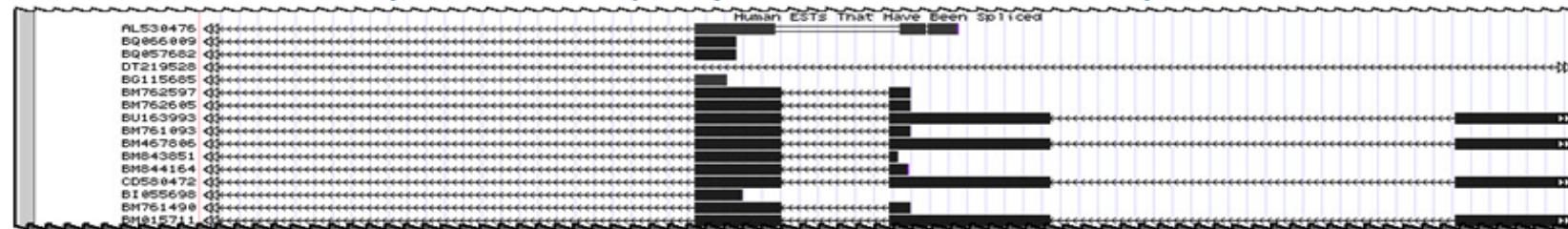
- Squish: each item = separate line, but 50% height + packed



- Pack: each item separate, but efficiently stacked (full height)



- Full: each item on separate line (may need to zoom to fit)



# Additional Options: Filters, Supertracks ...

The screenshot illustrates several tracks and their settings:

- Human ESTs Track Settings:** Shows a dropdown menu with "pack" selected. A red arrow points to the "Human ESTs Including Unspliced" section. The "Filter" dropdown is set to "red". A red circle highlights the "tissue: neuroblastoma" filter.
- ENCODE TF Binding Super-track Settings:** Shows a dropdown menu with "show" selected. A red arrow points to the "ENC TF Binding..." section. The "Display mode" dropdown is set to "dense". A red circle highlights the "Uniform TFBS" checkbox.
- Human ESTs Including Unspliced:** A detailed genomic track showing multiple ESTs across a genomic region. The track includes labels for Spliced ESTs, Human ESTs, and Unspliced ESTs. A red circle highlights a specific EST entry.
- Filter Selection:** A separate window titled "Select subtracks by cell line and factor". It lists factors on the left and cell lines on the top right. A red arrow points to the "ATF2" checkbox under the "ATF1" factor row. A red circle highlights the "ATF2" checkbox.

On      Off

- Some tracks have filters (*ESTs shown; SNPs other good example*)
- Super-tracks may have multiple components, various settings
- Some tracks may have un-displayed data

# Mid page options to change settings

The screenshot shows the UCSC Genome Browser interface with several configuration options highlighted:

- Top Control Bar:** Includes "move start" and "move end" buttons with a value of "2.0".
- Track Control Buttons:** Buttons for "track search", "default tracks", "default order", "hide all" (highlighted with a red box), "add custom tracks", "track hubs", "configure" (highlighted with a red box), "reverse", "resize", and "refresh".
- Configure Image Dialog:** A modal window titled "Configure Image" with fields for "image width: 1000 pixels", "label area width: 17 characters", and "text size: 8". It also contains three checked checkboxes:
  - Display chromosome ideogram above main graphic
  - Show light blue vertical guidelines
  - Display labels to the left of items in tracks
- Configure Tracks Dialog:** A modal window titled "Configure Tracks on UCSC Genome Browser: Human Feb. 2009 (GRCh37/hg19)". It lists tracks and their visibility status:

Track	Status	Description
Base Position	dense	Chromosome position in bases. (Clicks here zoom in 3x)
Chromosome Band	hide	Chromosome Bands Localized by FISH Mapping Clones
STS Markers	hide	STS Markers on Genetic (blue) and Radiation Hybrid (black) Maps
FISH Clones	hide	Clones Placed on Cytogenetic Map Using FISH
Recomb Rate	hide	Recombination Rate from deCODE, Marshfield, or Genethon Maps (deCODE default)
deCODE Recomb	hide	deCODE Recombination maps, 10Kb bin size, October 2010

- Search for data types
- Reset to defaults
- Configure options page
- You control the views with numerous features

# Printing track figures

- Customize track
- Add title
- consider showing only one transcript per gene by turning off splice variants
- Increase the font size and remove the light blue vertical guide lines in the image configuration menu
- Change image size
- Click on blue navigation menu-> view ->**PDF/PS link**

# Retrieve DNA sequence

 Genomes   Genome Browser   Tools   Mirrors   Downloads   My Data   Help   About Us

Get DNA in Window (hg19/Human)

blue navigation menu -> view-> DNA

## Get DNA for

Position

Note: This page retrieves genomic DNA for a single region. If you would prefer to get DNA for many items in a particular track, or get DNA with formatting options based on gene structure (introns, exons, UTRs, etc.), try using the [Table Browser](#) with the "sequence" output format.

### Sequence Retrieval Region Options:

Add  extra bases upstream (5') and  extra downstream (3')

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

### Sequence Formatting Options:

- All upper case.
- All lower case.
- Mask repeats:  to lower case  to N
- Reverse complement (get '-' strand sequence)

Note: The "Mask repeats" option applies only to "get DNA", not to "extended case/color options".

## 2. BLAT (Blast Like Alignment Tool)

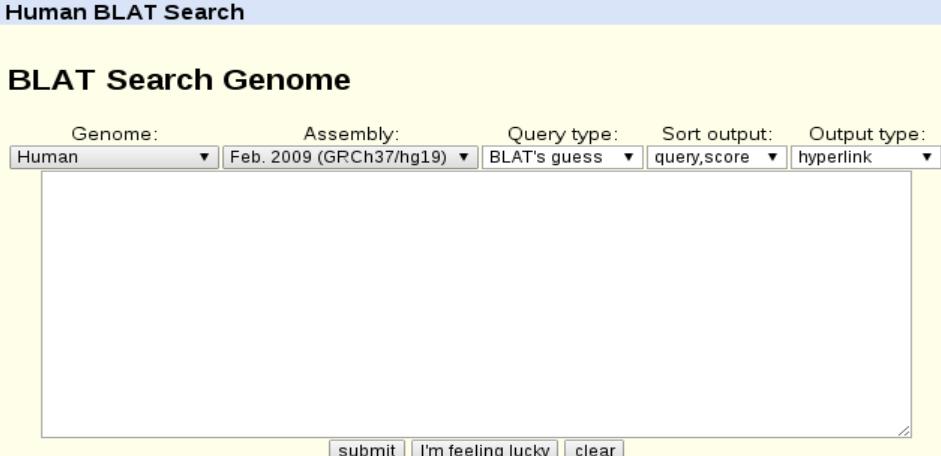
- Rapid sequence search by indexing entire genome
- Useful for finding high similarity matches
- 95% and greater similarity of length 25 bases or more OR sequences of 80% and greater similarity of length 20 amino acids or more
- Limits: DNA (25000 bp), Protein (10000 aa) or 25 sequences
- Can be installed and run locally

Human BLAT Search

BLAT Search Genome

Genome:	Assembly:	Query type:	Sort output:	Output type:
Human	Feb. 2009 (GRCh37/hg19)	BLAT's guess	query,score	hyperlink

submit I'm feeling lucky clear



# BLAT results

## Human BLAT Results

### BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN	
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	2581	1	2591	2591	100.0%	17	-	7571720	7590868	19149
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	177	2158	2436	2591	83.1%	1	+	45290354	45290634	281
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	176	2134	2433	2591	85.6%	10	-	27408468	27408791	324
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	174	2141	2437	2591	83.7%	2	+	27384674	27384975	302
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	174	2134	2436	2591	87.6%	10	+	67312526	67312836	311
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	173	2148	2431	2591	87.4%	10	+	71133346	71133631	286
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	173	2149	2504	2591	84.0%	10	+	65420577	65421000	424
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	172	2153	2433	2591	83.4%	3	+	27600067	27600347	281
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	165	2160	2444	2591	88.4%	X	-	122127686	122127972	287
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	162	2152	2435	2591	83.2%	2	-	109493652	109493934	283
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	162	2137	2434	2591	84.0%	1	-	225930110	225930396	287
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	162	2144	2437	2591	83.5%	10	+	15559328	15559614	287
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	160	2138	2552	2591	82.9%	9	-	131379044	131379531	488
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	160	2158	2435	2591	82.2%	4	-	139925816	139926096	281
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	160	2134	2414	2591	84.3%	10	-	12095247	12095528	282
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	160	2127	2434	2591	86.0%	2	+	170700494	170700797	304
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	153	2183	2435	2591	85.4%	1	-	106084596	106084833	237
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	26	2128	2154	2591	100.0%	3	-	27607611	27607638	28
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	26	2408	2437	2591	93.4%	X	+	47169213	47169242	30
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	26	2273	2304	2591	90.7%	5	+	7460469	7460500	32
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	25	2358	2389	2591	82.8%	2	+	124842060	124842089	30
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	23	2353	2379	2591	92.6%	X	-	100332288	100332314	27
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	23	2323	2345	2591	100.0%	X	+	47169722	47169744	23
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	22	2369	2404	2591	80.6%	20	-	33243008	33243043	36
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	21	2182	2202	2591	100.0%	2	+	38998603	38998623	21
<a href="#">browser</a>	<a href="#">details</a>	uc002gij.3	21	2347	2367	2591	100.0%	1	+	199938363	199938383	21

- Results with demo sequences, settings default; sort = Query, Score
  - **Score is a count of matches—higher number, better match**
- Click [browser](#) to go to Genome Browser image location (next slide)
- Click [details](#) to see the alignment to genomic sequence (2nd slide)

# Browser link

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

chr17:7,571,720-7,590,868 19,149 bp. enter position, gene symbol or search terms go

Scale 5 kb hg19

chr17 (p13.1) 15.3 15.2 15.1 17q11.2 17q11.2 17q12 21.31 21.32 21.2 24.2 24.1 24.0 24.1 17q25.3

query →

**YOUR Sequence from BLAT Search**

UCSC Genome Browser (Searched, Opened, Opened, Opened, Opened, Opened, Opened, Opened)

TPS3 TPS3 TPS3 TPS3 TPS3 TPS3 TPS3 TPS3

**Mapping and Sequencing Tracks**

refresh

Base Position	Chromosome Band	STS Markers	FISH Clones	Recomb Rate	deCODE Recomb
dense	hide	hide	hide	hide	hide

**ENCODE**

Pilot Map Contigs Assembly GRC Map Contigs Gap Publications

[No hide hide hide hide hide hide]

data-chr17 Fosmid End Pairs GC Percent GRC Patch Hg18 Diff GRC Incident

BAC End Pairs Pairs hide hide hide hide hide

Hi Seq Depth Wiki Track BU ORCHID Mapability Short Match Restr Enzymes

Blat Sequence pack

- From browser click in BLAT results
- A new track line with *Your Sequence from BLAT Search* appears

# Details link

Alignment of uc002gij.3

uc002gij.3

Human chr17

block1

block2

block3

block4

block5

block6

block7

block8

block9

block10

block11

together

Alignment of uc002gij.3 and chr17:7571720-7590868

Click on links in the frame to the left to navigate through the alignment. Matching bases in cDNA and genomic sequences are colored blue and capitalized. Light blue bases mark the boundaries of gaps in either sequence (often splice sites).

cDNA uc002gij.3

Your query

Genomic chr17 (reverse strand):

tccccaactc catttccctt gtttccctcg gcaggccgt tacttgccc 7590919  
tacttgtcat gggtggatgc cagggtgtc ccaggacgt cgcagggtt 7590869  
**GATGGGATTG GGGTTTCCC CTCCCATGTG CTCAGACTG GCGCTAAAG 7590819**  
TTTGAGCTT CTCRAAAGTC TAGAGCCACC GTCCAGGGAG CAGGAGCTG 100  
CTGGGCCTCG GGGACAC 7590819  
CGGGCACAGC CTTCCTCG 7590819  
CCATCCACCA GCGGCC 7590819  
GAGACATTT GAGACCA 7590819  
CCCCCTCCCG TCCCAA 7590819  
TTGAACAATG GTTCAC 7590819  
CCAGGAGCTG CTCCCTCG 7590819  
GGCCCTTGCA CGCCCG 7590819  
AGRAAACCTA CCAGGG 7590819  
GGGACAGCCA AGTCGTG 7590819  
GTTCAGGATTC CTGAGCC 7590819  
CACCCCGGGG COCCAC 7590819  
CAUCACATTA CCGAGG 7590819  
AGATAACCGAT GTCCTG 7590819  
ATTCGCTGAT GAGGAT 7590819  
GTGGTGCCTT ATGAGC 7590819  
CTACAACATC ATGTTG 7590819  
CCATCCCTCA CTGCTG 7590819  
CGGAAACGCT TGTGGG 7590819  
CACAGGAGGA GAGAGG 7590819  
CCCCACACAC CACTAC 7590819  
CAUCACAAAG AUAAAAC 7590819  
TGGGCTGTGAG CGCTCG 7590819  
TCAGAAGATTC CGCAAG 7590819  
AGCCACCTTA AGTCCTA 7590819  
CATGTCAGG ACAGAA 7590819

Genomic match, with color cues

Side by Side Alignment

yours genomic

Side by Side Alignment

your genomic

### 3. Custom tracks, session and track Hubs

#### Sessions

- **Signing in** enables you to save current settings into a named session, and then restore settings from the session later.
- lifespan: 4 months
- If you wish, you can share named sessions with other users.
- Individual sessions may be designated as either *shared* or *non-shared* to protect the privacy of confidential data.

The screenshot shows the UCSC Genome Bioinformatics session management interface. At the top, there is a navigation bar with links for Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data (which is highlighted with a red box), Help, and About Us. A dropdown menu under 'My Data' includes options for Sessions, Track Hubs, and Custom Tracks. Below the navigation bar, there is a 'Sign in to UCSC Genome Bioinformatics' section with links for Login and Create an account. A note states that signing in enables saving current settings into a named session and restoring them later, and that sessions can be shared with others. The main content area is titled 'Session Management'. It includes a note about the Sessions User's Guide, a link to reset browser settings, and information about saving sessions. There are sections for 'Save Settings' (with a form to save current settings to a local file) and 'Restore Settings' (with forms to use another user's saved session, a local file, or a URL). The bottom section is titled 'Sharing Sessions' with a note about sharing saved sessions with others, followed by two bullet points describing how to do this.

Sign in to UCSC Genome Bioinformatics

Login  
Create an account

Signing in enables you to save current settings into a named session, and then restore settings from the session later. If you wish, you can share named sessions with other users.

#### Session Management

See the [Sessions User's Guide](#) for more information about this tool.

[Click here to reset](#) the browser user interface settings to their defaults.

If you [sign in](#), you will also have the option to save named sessions.

#### Save Settings

Save current settings to a local file:

file:  file type returned:  submit  
(leave file blank to get output in browser window)

#### Restore Settings

Use settings from another user's saved session:

user:  session name:  submit

Use settings from a local file: Choose file No file chosen submit

Use settings from a URL (<http://...>, <ftp://...>):  submit

#### Sharing Sessions

There are several ways to share saved sessions with others.

- If you [sign in](#), you will be able to save named sessions which will be displayed with Browser and Email links.
- If you have saved your settings to a local file, you can send email to others with the file as an

# Custom tracks

it is possible for users to upload their own annotation data for temporary display in the browser. These custom annotation tracks are viewable only on the machine from which they were uploaded and are automatically discarded 48 hours after the last time they are accessed, unless they are saved in a [Session](#). Optionally, users can make custom annotations viewable by others as well.

- Format your data
- Define browser characteristics
- Define track characteristics
- Upload and view your track
- Add URL for annotation details (option)

# Track Hubs

Home Genomes Blat Tables Gene Sorter PCR Session FAQ Help

Human (*Homo sapiens*) Genome Browser Gateway

The UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#).  
Software Copyright (c) The Regents of the University of California. All rights reserved.

clade genome assembly position or search term gene image width  
Mammal Human Feb. 2009 (GRCh37/hg19) chr21:33,031,597-33,041,570 1134 submit

**Track Data Hubs**

Track data hubs are collections of tracks from outside of UCSC that can be imported into the Genome Browser. To import a public hub check the box next to the hub name. The hub will show up as a group of tracks with its own blue bar and label underneath the main browser graphic, and in the configure page. For more information see the [Hub Guide](#).

NOTE: Because Track Hubs are created and maintained by external sources, UCSC is not responsible for their content.

**UCSC Genome Browser on Human Mar. 2006 (NCBI36/hg18) Assembly**

move <<< << < > >> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x  
chr17:38,449,840-38,530,994 81,155 bp. enter position, gene symbol or search terms go

chr17 (q21.31) p13.3 q13.2 p13.1 17p12 17p11.2 17q11.2 17q12 17q11.3 17q22 23.2 24.2 q24.3 q25.1 17q25.3

Scale chr17: 38,460,000 38,470,000 38,480,000 38,490,000 38,500,000 38,510,000 38,520,000 38,530,000 hg18 Human\_CD133HSC\_Meth Changes in Human Hematopoietic Stem Cells, Hodges 2011

CD133HSC HSPC HSFC Neu Neu

**DNA Methylation**

refresh

Acute Myeloid Leukemia B Cells Blood Cells from Different Ages Brains Breast Cancer Chronic Lymphocytic Leukemia Colon Cancer and Adenomatous Polyp Developing human brain Fetal Lung Fibroblasts Fibroblasts Hematopoietic Stem Cells Induced Pluripotent Stem Cells Leukocytes Lymphoblastoid Neuroepithelium Cells Neuronal Cells Peripheral Blood Mononuclear Cells Sperm

Use Selected Hubs Load Dsoe.ucsc.

56

# Track Hubs

**Table Browser**

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser tutorial for a narrated presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

clade: Mammal   genome: Human   assembly: Mar. 2006 (NCBI36/hg18)  

group: Custom Tracks   track: clones   manage custom tracks   track hubs

table: clones\_7284   describe table schema

region:  genome  ENCODE Pilot regions  position chr4:56010000-56030000   lookup   define regions

Identifiers (names/acccessions):  
paste list   upload list

filter: [create](#)

Intersection with knownGene:  
correlation: [create](#)

output format: custom track

output file:

file type returned:  plain text  XML

Note: The all fields and selected fields  
[get output](#)   [summary/statistics](#)

To reset all user cart settings (including custom tracks), [click here](#).

**Track Hubs**

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser tutorial for a narrated presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data.

clade: Mammal   genome: Human   assembly: Feb. 2009 (GRCh37/hg19)  

group: [Custom Tracks](#)   track: [All Summary](#)   add custom tracks   track hubs

table: [Phenotype and Disease Associations](#)   describe table schema

region: [Genes and Gene Prediction Tracks](#)   lookup   define regions

filter: [Expression](#)   [Regulation](#)

Identify: [Comparative Genomics](#)

output: [Residential Assembly and Analysis](#)

output: [Variation and Repeats](#)

file type: [All Tracks](#)   [All Tables](#)

as:  GREAT  
keep output in browser

Note: to return more than 100,000 lines, change the filter setting (above). The entire data set may be available for download as a very large file that contains the original data values (not compressed into the wiggle format) – see the [Downloads](#) page.  
[get output](#)   [summary/statistics](#)

To reset all user cart settings (including custom tracks), [click here](#).

## 4. UCSC Table Browser

- Search for genes and annotation
- Setup and filters
- Join tables
- Retrieve sequences
- Intersecting tracks
- Export to external resources

# Table browser interface

clade: Mammal ▾ genome: Human ▾ assembly: Feb. 2009 (GRCh37/hg19) ▾  
group: Mapping and Sequencing ▾ track: Assembly ▾ add custom tracks track hubs  
table: gold ▾ describe table schema  
region:  genome  ENCODE Pilot regions  position chr19:313707-313990 lookup define regions  
identifiers (names/accessions): paste list upload list  
filter: create  
intersection: create  
correlation: create  
output format: all fields from selected table ▾ Send output to  Galaxy  GREAT  GenomeSpace  
output file: (leave blank to keep output in browser)  
file type returned:  plain text  gzip compressed

get output summary/statistics

To reset **all** user cart settings (including custom tracks), [click here](#).

# Table browser usage

- Retrieve the DNA [sequence data or annotation data](#) underlying Genome Browser tracks for the entire genome, a specified coordinate range, or a set of accessions
- Apply a [filter](#) to set constraints on field values included in the output
- Generate a [custom track](#) and automatically add it to your session so that it can be graphically displayed in the Genome Browser
- Conduct both structured and free-from SQL queries on the data
- Combine queries on multiple tables or custom tracks through an [intersection or union](#) and generate a single set of output data
- Display [basic statistics](#) calculated over a selected data set
- Display the schema for table and list all other tables in the database connected to the table
- Organize the [output data](#) into several different formats for use in other applications, spreadsheets, or databases

# Table Browser driven discovery

Task: Search entire genome for “CAG” trinucleotide repeats from UCSC tables.

- Choose genome [[hg19](#)]
- Choose table [[Repeats->Simple Repeats](#)]
- Describe table -find correct data fields
- Choose region [[genome](#)]
- Upload locations
- Data summary - approx. 1 million simple repeats

## Features of trinucleotide expansion in humans

Disease	Sequence	Location	Parent of origin of expansion	Repeat number (normal)	Repeat number (pre-mutation)	Repeat number (disease)	Somatic instability
<b>Diseases with coding TNRs</b>							
DRPLA	CAG	<i>ATN1</i> (exon 5)	P	6–35	35–48	49–88	Yes
HD	CAG	<i>HTT</i> (exon 1)	P	6–29	29–37	38–180	Yes
OPMD	GCN	<i>PABPN1</i> (exon 1)	P and M	10	12–17	>11	None found in tissue tested (hypothalamus)
SCA1	CAG	<i>ATXN1</i> (exon 8)	P	6–39	40	41–83	Yes
SCA2	CAG	<i>ATXN2</i> (exon 1)	P	<31	31–32	32–200	Unknown
SCA3 (Machado– Joseph disease)	CAG	<i>ATXN3</i> (exon 8)	P	12–40	41–85	52–86	Unknown
SCA6	CAG	<i>CACNA1A</i> (exon 47)	P	<18	19	20–33	None found
SCA7	CAG	<i>ATXN7</i> (exon 3)	P	4–17	28–33	>36 to >460	Yes
SCA17	CAG	<i>TBP</i> (exon 3)	P > M	25–42	43–48	45–66	Yes
SMBA	CAG	<i>AR</i> (exon 1)	P	13–31	32–39	40	None found

**McMurray CT. Mechanisms of trinucleotide repeat instability during human development. Nat Rev Genet. 2010 Nov;11(11):786-99.**

# Table Browser: Filtering

clade: Mammal   genome: Human   assembly: Feb.  
group: All Tracks   track: Common SNPs(141)  
table:.snp141Common   describe table schema  
region: genome   ENCODE Pilot regions   position chr19:313707-313  
identifiers (names/acceessions): paste list upload list  
filter: create  
intersection: create  
correlation: create create  
output format: all fields from selected table Send output to output file: (leave blank to keep output  
file type returned: plain text   gzip compressed

Simple Repeats (simpleRepeat) Summary Statistics

item count	80
item bases	3,222 (0.00%)
item total	3,222 (0.00%)
smallest item	25
average item	40
biggest item	93
smallest score	50
average score	67
biggest score	130

Results

Filter on Fields from table

AND

bin	is ignor	does match	AND
chrom	is does	match 0	AND
chromStart	is ignored ▾	0	AND
chromEnd	is ignored ▾	0	AND
name	does ▾	match *	AND
period	is ignored ▾	0	AND
copyNum	is ignored ▾	0	AND
consensusSize	is ignored ▾	0	AND
perMatch	is ignored ▾	0	AND
perIndel	is ignored ▾	0	AND
score	is ignored ▾	0	AND
A	is ignored ▾	0	AND
C	is ignored ▾	0	AND
G	is ignored ▾	0	AND
T	is ignored ▾	0	AND
entropy	is ignored ▾	0	AND
sequence	does ▾	match CAG	AND

AND ▾ Free-form query:

search for simple repeats in the entire genome with “CAG” sequence and extract data table.

# Table Browser: Intersections

- Combines the output of two queries into a single set of data based on specific join criteria.
- For example, this can be used to find all SNPs that intersect with RefSeq coding regions. The intersection can be configured to retain the existing alignment structure of the table with a specified amount of overlap, or discard the structure in favor of a simple list of position ranges using a base-pair intersection or union of the two data sets.
- The button functionalities are similar to those of the *filter* option.

## Other tools

- Gene sorter
- *In silico* PCR
- VisiGene browser
- Cancer Browser and Encode portal
- Genome graphs
- Other tools:
  - liftOver
  - Dusters
  - Tree maker

# Search for related genes

## UCSC Human Gene Sorter

genome  assembly  search  Go!  
sort by  configure filter (now off) display  output

### About the Gene Sorter

This program displays a sorted table of genes that are related to one another. The relationship can be one of several types, including protein-level homology, similarity of gene expression profiles, or genomic proximity.

To display a gene and its relatives:

1. Select a genome and assembly from the corresponding pull-down menus.
2. Type a word or phrase into the *search* text box to specify which gene should be displayed in the Gene Sorter. Examples of search terms include FOXA2, HOXA9, and MAP kinase.
3. Choose the gene relationship with which you would like to sort the list by selecting an option from the *sort by* pull-down menu.
4. Press the *Go!* button to display your results.

Following a successful search, the Gene Sorter displays a table containing the specified gene -- highlighted in light green -- and its relatives, each on a separate line. To adjust the number of rows shown, select an option from the *display* pull-down menu.

The default set of table columns -- which can be expanded, reduced, and rearranged via the *configure* button -- shows additional information about the genes. Some of the column data, such as those in the *BLAST E-value* and *%ID* columns, are calculated relative to the highlighted gene. To select a different gene in the list, click on its name. Clicking on a gene's *Genome Position* will open the UCSC Genome Browser to the location of that gene. Similarly, clicking on a gene's *Description* will open a page showing detailed information about the gene.

One of the most powerful features of the Gene Sorter is its filtering capabilities, accessed via the *filter* button. Use the filter to fine-tune the list of displayed genes to a subset based on a selection of detailed and flexible criteria. For example, the filter may be used to select all human genes over-expressed in the cerebellum that have GO-annotated G-protein coupled receptor activity.

The Gene Sorter offers two options for displaying and downloading sequence associated with the genes in the table. Clicking on the *sequence* button will fetch associated protein, mRNA, promoter, or genomic sequence. To dump the table into a simple tab-delimited format suitable for import into a spreadsheet or relational database, click the *text* button.

The UCSC Gene Sorter was designed and implemented by Jim Kent, Fan Hsu, Donna Karolchik, David Haussler, and the UCSC Genome Bioinformatics Group. This work is supported by a grant from the National Human Genome Research Institute and by the Howard Hughes Medical Institute.

# Gene Sorter

## UCSC Human Gene Sorter

genome Human ▾ assembly Mar. 2006 (NCBI36/hg18) ▾ search uc002gij.2  
 sort by Expression (GNF Atlas2) ▾ configure filter (now off) display 25 ▾ output sequence text ▾ Go!

#	Name	VisiGene	fetal brain	whole brain	amygdala	thymus	bone marrow	PB-CD4+ Tcells	skin	adipocyte	pancreatic islets	heart	lung	kidney	liver	ovary	testis	BLASTP E-Value	Genome Position	Description
1	TP53	n/a																0	chr17 7,522,016	tumor protein p53 isoform a
2	RPS20	n/a																n/a	chr8 57,148,895	ribosomal protein S20
3	H2AFV	n/a																n/a	chr7 44,846,994	H2A histone family, member V isoform 1
4	RPL7A	187765																n/a	chr9 135,206,495	ribosomal protein L7a
5	RPS13	n/a																n/a	chr11 17,054,155	ribosomal protein S13
6	SNRPG	181122																n/a	chr2 70,368,191	small nuclear ribonucleoprotein polypeptide G
7	EIF4A1	176036																n/a	chr17 7,419,687	eukaryotic translation initiation factor 4A
8	ADSL	77625																n/a	chr22 39,082,485	adenylosuccinate lyase isoform a
9	CR601950	n/a																n/a	chr17 72,069,204	Homo sapiens primary hepatoblastoma cDNA, clone:HKMT0728, full insert sequence.
10	UBE2A	182203																n/a	chrX 118,597,467	ubiquitin-conjugating enzyme E2A isoform 1
11	GMPS	176663																n/a	chr3 157,104,616	guanine monophosphate synthetase
12	G3BP1	176455																n/a	chr5 151,148,388	Ras-GTPase-activating protein SH3-domain-binding
13	NUP37	187198																n/a	chr12 101,014,297	nucleoporin 37kDa
14	QARS	180161																n/a	chr3 49,112,772	glutaminyl-tRNA synthetase
15	ZNF207	26352																n/a	chr17 27,711,425	zinc finger protein 207 isoform c
16	XRCC5	n/a																n/a	chr2 216,730,812	ATP-dependent DNA helicase II
17	LOC647099	n/a																n/a	chr17 24,073,314	similar to ribosomal protein L23A
18	PABPC4	36799																n/a	chr1 39,807,039	poly A binding protein, cytoplasmic 4 isoform 2
19	RPS18	180521																n/a	chr6 33,350,044	ribosomal protein S18
20	RPS18	n/a																n/a	chr6_cox_hap1 4,622,203	ribosomal protein S18
21	RPS18	n/a																n/a	chr6_qbl_hap2 4,428,251	ribosomal protein S18
22	PSMA5	180067																n/a	chr1 109,758,277	proteasome alpha 5 subunit
23	LOC441743	n/a																n/a	chr16 376,999	Uncharacterized protein ENSP00000332117.
24	PHF10	27218																n/a	chr6 169,855,917	PHD finger protein 10 isoform a
25	RPS27	59894																n/a	chr1 152,230,551	ribosomal protein S27

# Configure

## Configure Gene Sorter

Columns:    Settings:    
 Expression ratio colors:    Show all splicing variants:

Name	On	Position	Description	Configuration
#	<input checked="" type="checkbox"/>		Item Number in Displayed List/Select Gene	n/a
Name	<input checked="" type="checkbox"/>		Gene Name/Select Gene	n/a
UniProtKB	<input type="checkbox"/>		UniProtKB Protein Display ID	n/a
UniProtKB Acc	<input type="checkbox"/>		UniProtKB Protein Accession	n/a
RefSeq	<input type="checkbox"/>		NCBI RefSeq Gene Accession	n/a
Entrez Gene	<input type="checkbox"/>		NCBI Entrez Gene/LocusLink ID	n/a
UCSC ID	<input type="checkbox"/>		UCSC Transcript ID	n/a
GenBank	<input type="checkbox"/>		GenBank mRNA Accession	n/a
Ensembl	<input type="checkbox"/>		Ensembl Transcript ID	n/a
KEGG	<input type="checkbox"/>		KEGG Pathway ID	n/a
GNF Atlas 2 ID	<input type="checkbox"/>		ID of Associated GNF Atlas 2 Expression Data	n/a
Gene Category	<input type="checkbox"/>		High Level Gene Category - Coding, Antisense, etc.	n/a
CDS Score	<input type="checkbox"/>		Coding potential score from txCdsPredict	n/a
VisiGene	<input checked="" type="checkbox"/>		UCSC VisiGene In Situ Image Browser	n/a
Allen Brain	<input type="checkbox"/>		Allen Brain Atlas In Situ Images of Adult Mouse Brains	n/a
U133 ID	<input type="checkbox"/>		ID of Associated Affymetrix U133 Expression Data	n/a
U133Plus2 ID	<input type="checkbox"/>		ID of Associated Affymetrix U133 Plus 2.0 Expression Data	n/a
U95 ID	<input type="checkbox"/>		ID of Associated Affymetrix U95 Expression Data	n/a
GNF Atlas 2	<input checked="" type="checkbox"/>		GNF Expression Atlas 2 Data from U133A and GNF1H Chips	brightness: <input type="text" value="1.0"/> tissues: <input type="button" value="selected"/> values: <input type="button" value="ratio"/>
H-Inv	<input type="checkbox"/>		H-Invitational Gene Database	n/a
Max GNF Atlas 2	<input type="checkbox"/>		Maximum Expression Value of GNF Expression Atlas 2	n/a
GNF Atlas 2 Delta	<input type="checkbox"/>		Normalized Difference in GNF Expression Atlas 2 from Selected Gene	n/a
GNF U95	<input type="checkbox"/>		GNF Expression Atlas 1 Human Data on Affy U95 Chips	brightness: <input type="text" value="1.0"/> tissues: <input type="button" value="selected"/> values: <input type="button" value="ratio"/>
Max GNF U95	<input type="checkbox"/>		Maximum Expression Value of GNF Expression Atlas 1	n/a
GNF Atlas1 Delta	<input type="checkbox"/>		Normalized Difference in GNF Atlas 1 Expression from Selected Gene	n/a
Affy Exons	<input type="checkbox"/>		Affymetrix All Exon Microarrays	brightness: <input type="text" value="1.0"/>
Affy Exon Dst	<input type="checkbox"/>		Affymetrix All Exon Microarrays Distance	n/a
BLASTP Bits	<input type="checkbox"/>		NCBI BLASTP Bit Score	n/a
BLASTP	<input type="checkbox"/>		NCBI BLASTP E Value	n/a

# Filter

## Gene Sorter Filter

On this page you can restrict which genes appear in the main table based on the values in any column. Click the *submit* button to return to the main Gene Sorter page with the current filter settings applied.

Quickly obtain a list of gene names that pass the filter:

### Filter Controls for Displayed Columns:

#### Name - Gene Name/Select Gene

Name search (including \* and ? wildcards):   
Include if  words in search term match.  
Limit to items (no wildcards) in list:

#### VisiGene - UCSC VisiGene In Situ Image Browser

VisiGene search (including \* and ? wildcards):   
Include if  words in search term match.  
Limit to items (no wildcards) in list:

#### GNF Atlas 2 - GNF Expression Atlas 2 Data from U133A and GNF1H Chips

Note: the values here range from about -5.0 to 5.0.  
These are calculated as logBase2(tissue/reference).

Tissue	Minimum	Maximum
fetal brain		
whole brain		
amygdala		
thymus		
bone marrow		
PB-CD4+ Tcells		
skin		
adipocyte		
pancreatic islets		
heart		
lung		
kidney		

# In silico PCR

## UCSC In-Silico PCR

Genome:	Assembly:	Target:	Forward Primer:	Reverse Primer:	<input type="button" value="submit"/>
<input type="button" value="Human"/>	<input type="button" value="Mar. 2006 (NCBI36/hg18)"/>	<input type="button" value="genome assembly"/>	<input type="text"/>	<input type="text"/>	
Max Product Size: <input type="text" value="4000"/>	Min Perfect Match: <input type="text" value="15"/>	Min Good Match: <input type="text" value="15"/>	Flip Reverse Primer: <input type="checkbox"/>		

## About In-Silico PCR

In-Silico PCR searches a sequence database with a pair of PCR primers, using an indexing strategy for fast performance.

### Configuration Options

**Genome and Assembly** - The sequence database to search.

**Target** - If available, choose to query transcribed sequences.

**Forward Primer** - Must be at least 15 bases in length.

**Reverse Primer** - On the opposite strand from the forward primer. Minimum length of 15 bases.

**Max Product Size** - Maximum size of amplified region.

**Min Perfect Match** - Number of bases that match exactly on 3' end of primers. Minimum match size is 15.

**Min Good Match** - Number of bases on 3' end of primers where at least 2 out of 3 bases match.

**Flip Reverse Primer** - Invert the sequence order of the reverse primer and complement it.

### Output

When successful, the search returns a sequence output file in fasta format containing all sequence in the database that lie between and include the primer pair. The fasta header describes the region in the database and the primers. The fasta body is capitalized in areas where the primer sequence matches the database sequence and in lower-case elsewhere. Here is an example from human:

```
>chr22:31000551+31001000 TAACAGATTGATGTCATGAAATGG CCCATGAGTGGCTCTAAAGCAGCTGC  
TtACAGATTGATGTCATGAAATGGggggggccagggggggggtga  
gactcgagaaaggcaggctggttcataacaacgttttgccgtccaa  
tatgcacgtcaaggtttcagggtctatgttgagccagtgggttaa  
tacacagaaatccatcgaaaaaccctatccattaaagataaaaaaa  
gacttgctgttgcataagggttttttttttttttttttttttttt  
tccagaatggcttaccccccacatgtcgaaaatgtgttacccgttaatctcaa  
agaacgtccctccatcgacagaaacaccacgcgtacaggaaacaaag  
aaattggcttcacttttaagggtgaaatccacaaaccaggatgtcgagactcc  
aaggactttgcgtcacCGACTGCTTAGAGCCACTCATGAG
```

The + between the coordinates in the fasta header indicates this is on the positive strand.

### Author

In-Silico PCR was written by [Jim Kent](#). Interactive use on this web server is free to all. Sources and executables to run batch jobs on your own server are available free for academic, personal, and non-profit purposes. Non-exclusive commercial licenses are also available. Contact Jim for details.

# *In silico* PCR usage

- Select genome
- Genomic or transcript?
- Enter primers
- Set configuration options

## About In-Silico PCR

In-Silico PCR searches a sequence database with a pair of PCR primers, using an indexing strategy for fast performance.

## Configuration Options

**Genome and Assembly** - The sequence database to search.

**Target** - If available, choose to query transcribed sequences.

**Forward Primer** - Must be at least 15 bases in length.

**Reverse Primer** - On the opposite strand from the forward primer. Minimum length of 15 bases.

**Max Product Size** - Maximum size of amplified region.

**Min Perfect Match** - Number of bases that match exactly on 3' end of primers. Minimum match size is 15.

**Min Good Match** - Number of bases on 3' end of primers where at least 2 out of 3 bases match.

**Flip Reverse Primer** - Invert the sequence order of the reverse primer and complement it.

## Output

When successful, the search returns a sequence output file in fasta format containing all sequence in the database that lie between and include the primer pair. The fasta header describes the region in the database and the primers. The fasta body is capitalized in areas where the primer sequence matches the database sequence and in lower-case elsewhere. Here is an example from human:

```
>chr22:31001000_31001000_TAACAGATTGATGATGAAATGGG CCCATGAGTGGCTTAAGCAGCTGC  
TACAGATTGATGATGAAATGGGgggtggccagggggggggggggg  
gactcgagaaaaaggcgggcgttgcataacaacgctttgttgtccaa  
tatgcacgtcgaaatttccaggggctgtatgttgatgcgcgtgggttaag  
tacacagacatcccataaaaaaccctcatcttcataaaaatataaa  
gacttgcgttgtaaaggattggattatccattttgagaaaaattctgtta  
tcagaatggcttacccacaatgtcgaaaaatgtgtaccgttaatctcaa  
agcaagctccctcagacagaaaaacaccacgcgtcacagaaqcaaaag  
aaattgcgttcactttaaagggtgaatttcagaaacctcgatgtcaagtc  
aaggactttgcgttcacgtccacGCCAGCTCTTTAGGAOCCTCATGAG
```

The + between the coordinates in the fasta header indicates this is on the positive strand.

# Visigene

## VisiGene Image Browser

VisiGene is a virtual microscope for viewing *in situ* images. These images show where a gene is used in an organism, sometimes down to cellular resolution. With VisiGene users can retrieve images that meet specific search criteria, then interactively zoom and scroll across the collection.

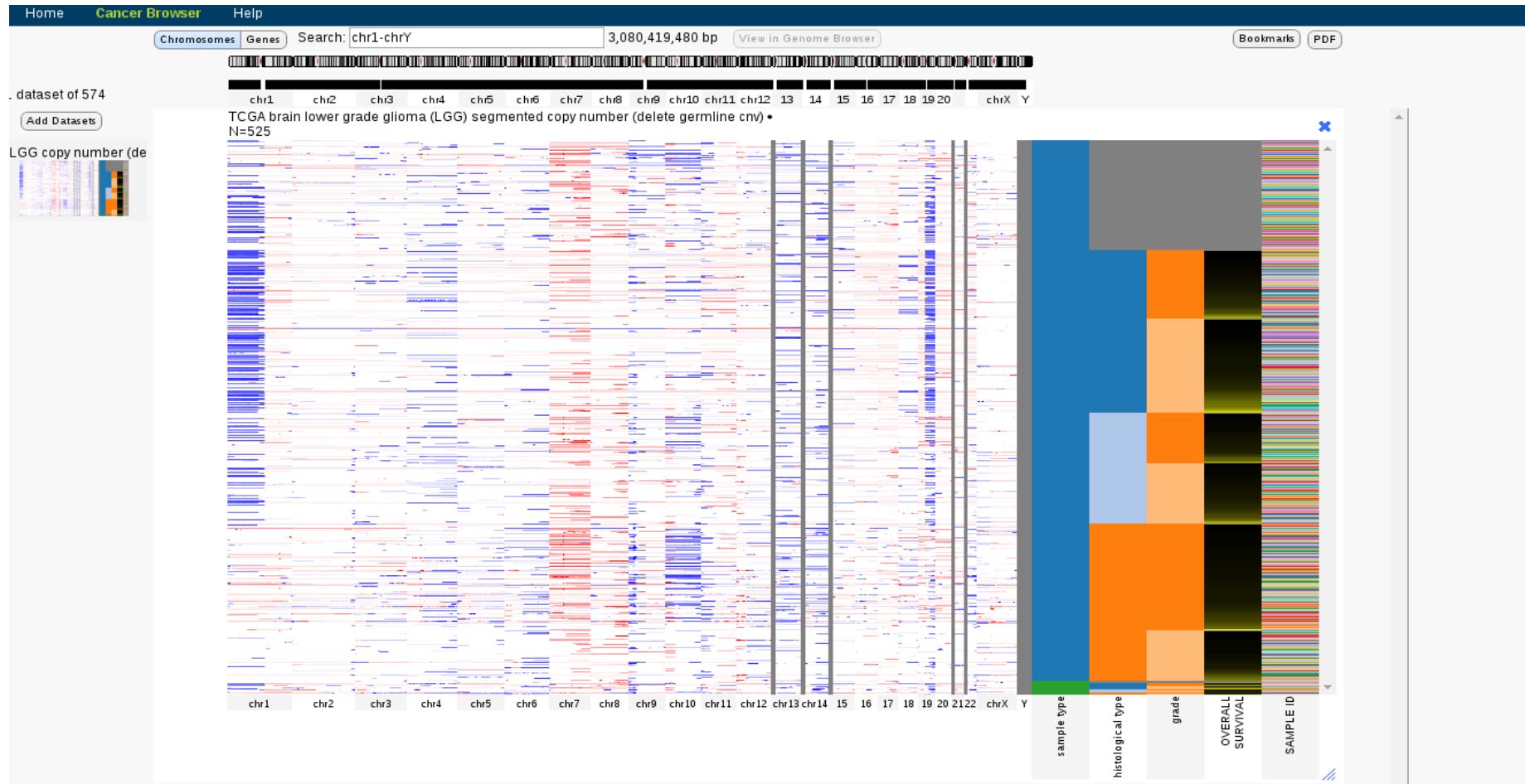
## Images Available

The following image collections are currently available for browsing:

- High-quality high-resolution images of eight-week-old male mouse sagittal brain slices with reverse-complemented mRNA hybridization probes from the [Allen Brain Atlas](#), courtesy of the [Allen Institute for Brain Science](#)
- Mouse *in situ* images from the [Jackson Lab Gene Expression Database](#) (GXD) at MGI
- Transcription factors in mouse embryos from the Mahoney Center for Neuro-Oncology
- Mouse head and brain *in situ* images from NCBI's [Gene Expression Nervous System Atlas](#) (GENSAT) database
- *Xenopus laevis* *in situ* images from the [National Institute for Basic Biology](#) (NIBB) XDB project



# Cancer Browser



# Encode



## Encyclopedia of DNA Elements at UCSC 2003 - 2012

**Human Data at UCSC**

- Downloads
- Experiment Matrix
- Search
- Genome Browser (hg19)
- Experiment List
- Cell Types
- Mouse Data at UCSC**
- Downloads
- Experiment Matrix
- Search
- Genome Browser (mm9)
- Experiment List
- Cell Types
- Metadata Terms**
- Registered Variables
- Antibodies
- Other Resources**
- News Archive
- First Production (2007-2012)
- Pilot (2003-2007)
- Contacts

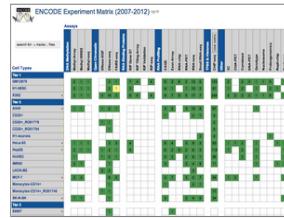
### About

The [Encyclopedia of DNA Elements](#) (ENCODE) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute ([NHGRI](#)). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.

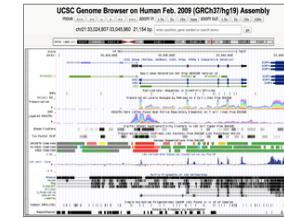
UCSC coordinated data for the ENCODE Consortium from its inception in 2003 (Pilot phase) to the end of the first 5 year phase of whole-genome data production in 2012. All data produced by ENCODE investigators and the results of ENCODE analysis projects from this period are hosted in the UCSC Genome browser and database. Explore ENCODE data using the image links below or via the left menu bar. **All ENCODE data at UCSC are freely available for download and analysis.**

**ENCODE results from 2013 and later are available from the ENCODE Project Portal,** [encodeproject.org](#). The ENCODE Project Portal also hosts ENCODE data from the first production phase, additional ENCODE access tools, and ENCODE project pages including up-to-date information about data releases, publications, and upcoming tutorials.

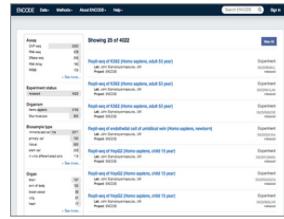
### Explore ENCODE data at UCSC



### View ENCODE data in the UCSC Genome Browser



### Search for data at the ENCODE Portal



### Search for ENCODE tracks in the UCSC Browser



# Other utilities

## UCSC Genome Bioinformatics

[Home](#) · [Genomes](#) · [Blat](#) · [Tables](#) · [Gene Sorter](#) · [PCR](#) · [Session](#) · [FAQ](#) · [Help](#)

### UCSC Genome Browser Utilities

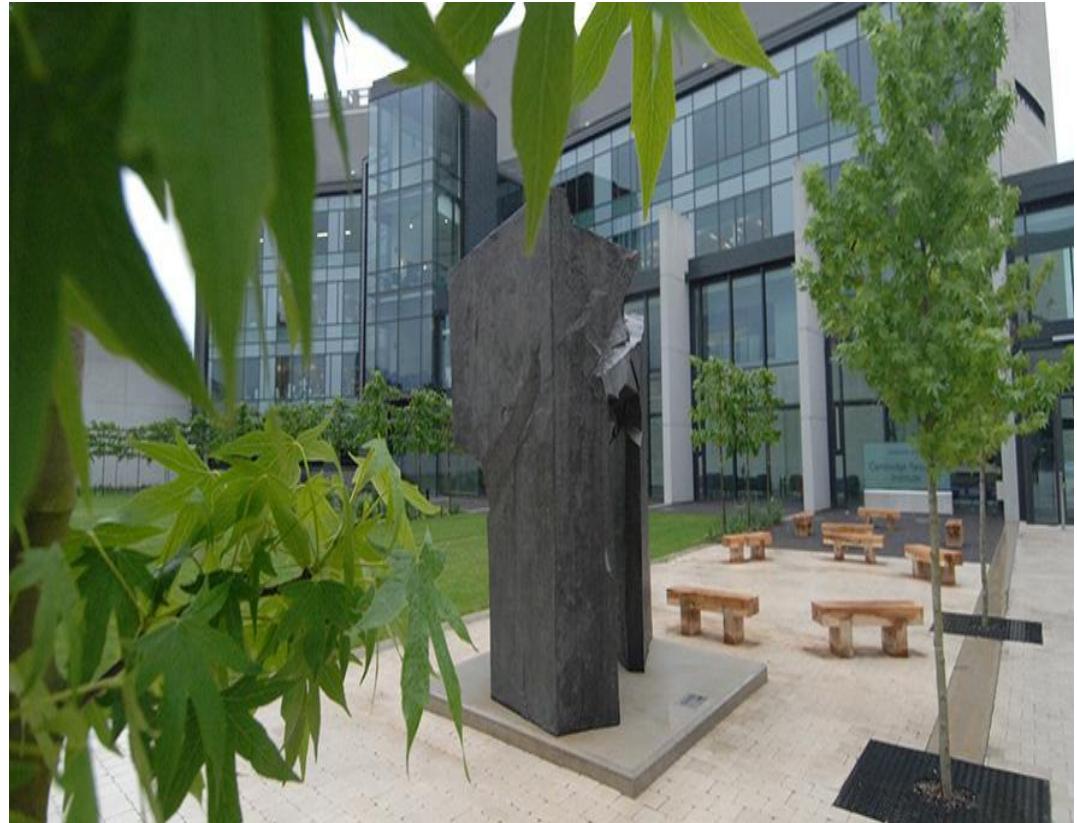
This page contains links to tools and utilities created by the UCSC Genome Bioinformatics Group.

- [Batch Coordinate Conversion \(liftOver\)](#) - converts genome coordinates and genome annotation files between assemblies. The current version supports both forward and reverse conversions, as well as conversions between selected species.
- [DNA Duster](#) - removes formatting characters and other non-sequence-related characters from an input sequence. Offers several configuration options for the output format, including translated protein.
- [Protein Duster](#) - removes formatting characters and other non-sequence-related characters from an input sequence. Offers several configuration options for the output format.
- [Phylogenetic Tree Gif Maker](#) - creates a gif image from the phylogenetic tree specification given. Offers several configuration options for branch lengths, normalized lengths, branch labels, legend etc.
- [Executable and Source Code Downloads](#) - executable and source code downloads of the Genome Browser, Blat and liftOver.

# Acknowledgements

CRUK CI

MRC Cancer Unit



Some slides were modified from UCSC and OpenHelix course material.